



Onderzoekers die dit artikel lezen, lezen ook...

Bogers, Toine; Van den Bosch, Antal

Published in:
DIXIT - Tijdschrift over taal- en spraaktechnologie

Publication date:
2011

Document version
Også kaldet Forlagets PDF

Citation for published version (APA):
Bogers, T., & Van den Bosch, A. (2011). Onderzoekers die dit artikel lezen, lezen ook... *DIXIT - Tijdschrift over taal- en spraaktechnologie*, 8(December), 42.

Onderzoekers die dit artikel lezen, lezen ook...

Samen met CLST van de Radboud Universiteit Nijmegen voerde de ILK onderzoeksgroep van het Tilburg center for Cognition and Communication (Tilburg University) van 2004 tot en met 2008 het À Propos-project uit. À Propos had tot doel kenniswerkers te ondersteunen in het professionele schrijfproces door relevante leesstof te suggereren tijdens het schrijfproces. Een specifiek voorbeeld is het vinden en suggereren van relevante artikelen voor wetenschappers die bezig zijn met het beschrijven van hun eigen onderzoek.

Toine Bogers
Royal School of
Library and
Information
Science,
Copenhagen,
Denmark
Antal van den
Bosch
CLST Radboud
Universiteit
Nijmegen

Deze pro-actieve automatische suggestie-generator is een voorbeeld van een zogenaamd recommender system. Een recommender system probeert de meest interessante en relevante objecten (zoals wetenschappelijke artikelen) in een collectie te vinden voor een specifieke gebruiker, of een groep gebruikers. Voor het genereren van zulke suggesties kunnen verschillende informatiebronnen gebruikt worden: demografische informatie, meta-data behorende bij de objecten, informatie over gebruikersvoorkeuren in het verleden, etc. Binnen het À Propos-project is gewerkt aan verschillende recommender systems: voor wetenschappelijke literatuur, voor nieuwsartikelen en voor het suggereren van experts.

Als onderdeel van het IOP-MMI-programma (een onderzoekssubsidieprogramma van Agent-schap NL, toen SenterNovem) was het opzetten van samenwerkingen met het bedrijfsleven voor À Propos van groot belang. Een voorbeeld van de praktische toepassing van het onderzoek verricht binnen À Propos, was de samenwerking met CiteULike, een Britse start-up. CiteULike is een zgn. social reference manager, een website die het mogelijk maakt voor haar gebruikers om hun favoriete wetenschappelijke artikelen online op te slaan en te beheren middels een web-interface, in plaats van deze artikelen lokaal op de computer op te slaan. Het 'sociale' aan het onderliggende systeem is dat het alle opgeslagen artikelen toegankelijk maakt voor alle gebruikers. CiteULike-gebruikers hebben de mogelijkheid om de artikelen die ze aan hun profiel hebben toegevoegd te beschrijven met steekwoorden. Deze steekwoorden worden ook wel tags genoemd en vormen een aanvulling bovenop de gebruikelijke metadata zoals titel en samenvatting. Tags, zoals bijvoorbeeld text mining of factorization, kunnen gebruikt worden om de artikelen van een gebruiker te categoriseren en verhogen de toegang tot en de hervindbaarheid van de artikelen. Tags worden beschikbaar gemaakt voor alle gebruikers. Veel gebruikers voegen deels dezelfde artikelen toe, met deels overlap-pende tags. Dit leidt tot het ontstaan van een

rijk netwerk van gebruikers, artikelen en tags, ook wel een folksonomy genoemd.

Voor het vinden van nieuwe, interessante artikelen moet een gebruiker in de klassieke zoekmachine-wereld zelf actie ondernemen: zoektermen intypen, browsen. Om dit proces makkelijker te maken ontwikkelde À Propos-medewerker Toine Bogers in samenwerking met CiteULike een recommender systeem voor wetenschappelijke artikelen in CiteULike. Bogers vergeleek het nut van de verschillende informatiebronnen, zoals de artikelen die door haar gebruikers toegevoegd zijn, metadata en tags. Hieruit bleek dat tags en metadata de beste mogelijkheden bieden om gelijksoortige, interessante artikelen te vinden voor een gebruiker. Artikelen die overeenkomen in de tags en metadata die eraan zijn toegekend leiden tot betere suggesties dan het gebruiken van informatie over gebruikersvoorkeuren in het verleden. De beste suggesties komen voort uit de combinatie van de verschillende algoritmes, zodat de zwaktes van de afzonderlijke algoritmes door elkaar worden opgevangen.

Dit recommender system voor wetenschappelijke artikelen werd in september 2010 toegevoegd op de CiteULike website. De initiële respons op de artikelsuggesties was erg positief, met een gemiddeld acceptatiepercentage van boven de 30%. Dit betekent dat van elke 10 suggesties die het recommender system maakt, er gemiddeld meer dan 3 toe worden gevoegd aan de collectie met artikelen die gebruikers bijhouden op CiteULike. In de toekomst zouden combinaties van meerdere recommendation-algoritmes dit acceptatiepercentage nog verder omhoog moeten kunnen stuwten, zoals al is aangetoond in experimenten die Bogers beschreef in zijn proefschrift.

Links

- <http://ilk.uvt.nl/apropos>
- www.citeulike.org
- <http://ilk.uvt.nl/~toine/phd-thesis>
- www.iva.dk

Figuur 1: Het À Propos recommender system in actie binnen de social bookmarking website CiteULike. De gebruiker wordt gewezen op een aantal artikelen die hij zelf nog niet gebookmarked heeft en die goed aansluiten bij wat hij al wel in CiteULike heeft ingebracht.

